



Timing matters: exploring the impact of pre- and post-draft rubrics on students' written assignments

Anastasiya A. Lipnevich¹ · Ligia Tomazin² · Samuel P. León³

Received: 30 November 2025 / Accepted: 15 September 2026
© The Author(s), under exclusive licence to Springer Nature B.V. 2026

Abstract

Rubrics are widely used in education to clarify expectations and guide student performance. However, the optimal timing of rubric presentation — whether before, after, or at both stages of a task — remains underexplored. This pre-registered experimental study aimed to determine whether rubrics function more effectively as pre-task guides or post-task feedback tools by examining differences in writing performance and cognitive load. Additionally, it explored potential differences between sixth and eleventh graders. Participants included 271 students from a private school in the Northeastern United States, with 141 eleventh graders and 130 sixth graders. Students wrote persuasive essays under one of four conditions: rubric before, rubric after, rubric before and after, or no rubric. Performance was assessed at two time points: the initial draft (T1) and the revised draft (T2). Cognitive load (perceived difficulty and effort) was also measured. Results revealed that rubrics provided before did not have any effect on the initial draft quality (T1), while rubrics provided after or at both stages enhanced revision outcomes (T2). Eleventh graders showed greater improvement than sixth graders. Cognitive load varied across conditions, with earlier rubric access reducing perceived difficulty. Thus, the timing of rubric presentation significantly influences writing performance and cognitive engagement. Rubrics are most effective when provided after the task, serving as an effective post-task feedback tool.

Keywords Rubrics · Cognitive load · Feedback · Writing task

✉ Anastasiya A. Lipnevich
a.lipnevich@gmail.com

¹ NBME, USA and The Graduate Center, City University of New York, New York, USA

² University of Trento, Italy and The Graduate Center, City University of New York, New York, USA

³ Universidad de Jaén, Jaen, Spain

Introduction

A rubric serves as a “coherent set of criteria for student’s work that includes descriptions of levels of performance quality on the criteria” (Brookhart, 2013, p. 4). This instructional tool is valuable because it can clearly delineate both evaluation criteria and performance standards (Andrade, 2000; English et al., 2022) as well as learning goals (Andrade, 2023). Both instructors and students have reported positive outcomes from implementing rubrics in classroom assessments (Atkinson & Lim, 2013; Reynders et al., 2020). Through their detailed articulation of expectations and learning objectives across multiple dimensions of a task, rubrics enable students to better understand the components of quality work. This enhanced understanding promotes students’ self-regulation (Panadero & Romero, 2014), boosts their self-efficacy (Lerdpornkulrat et al., 2019), and improves their overall learning and performance (Brookhart & Chen, 2015).

Moreover, rubrics increase transparency in assessment practices by explicitly outlining evaluation criteria (Brookhart, 2018). From an instructor’s perspective, these tools help maintain grading consistency by standardizing evaluations and reducing subjectivity (Chowdhury, 2018; Jonsson & Svingby, 2007). They also facilitate structured feedback that effectively identifies both strengths and areas for improvement. Research suggests that rubrics can streamline the grading process, potentially reducing the time instructors spend evaluating assignments (Allen & Tanner, 2006; Atkinson & Lim, 2013). Furthermore, rubrics play a crucial role in supporting formative assessment (Ghaffar et al., 2020) and have demonstrated significant potential in enhancing students’ self-assessment practices (Cockett & Jackson, 2018).

Research on rubric use in educational settings has increased significantly over recent decades (Dawson, 2017). Literature reviews consistently demonstrate rubrics’ positive impact on learning and instruction, highlighting their role in enabling formative feedback practices and supporting both teachers and students (Andrade, 2023; Brookhart & Chen, 2015; English et al., 2022; Jonsson & Svingby, 2007). These reviews also document improvements in students’ self-assessment abilities, self-regulation, and understanding of assessment criteria (Cockett & Jackson, 2018). Additionally, meta-analyses have provided quantitative evidence supporting rubrics’ positive effects on students’ self-efficacy, self-regulation, and academic performance (Panadero et al., 2023; Camargo Salamanca et al., 2024).

Although research consistently supports the benefits of rubric use for both teaching practices and student performance, one crucial aspect remains understudied: how the timing of rubric presentation affects the revision process and final outcomes. Rubrics are known to provide clear criteria for student performance, offering explicit expectations that guide both instructors and learners (Jonsson & Svingby, 2007). They help students understand the standards by which their work will be evaluated, promoting more focused and effective engagement with the assignment. Additionally, rubrics serve as valuable tools for self-assessment, allowing students to reflect on their work and identify areas for improvement (Lipnevich & Smith, 2022). However, despite the established advantages of rubrics, the question of whether providing rubrics before or after students produce an initial draft influences their improvements in their work has not been thoroughly explored. With the exception of Zhang et al. (2019), who examined this issue in a flipped learning context with undergraduate students, the timing of rubric presentation has received limited empirical attention. Thus, it

remains unclear whether students benefit more from seeing the rubric prior to drafting their work, allowing them to tailor their efforts to meet the specified criteria, or after submitting their drafts, as a tool for reflection, self-feedback generation, and revision. Understanding this distinction is crucial for optimizing rubric use to maximize student learning outcomes and improve the overall revision process.

With this experimental study, we aimed to address this question by investigating three distinct timing conditions for rubric presentation: before the initial draft, after its completion, or at both stages (i.e., before and after). We examined how these different timing conditions affect student performance on the final draft and explored the potential cognitive mechanism behind these differences. Through this investigation, we sought to determine whether rubrics serve more effectively as pre-writing guides or as feedback tools for revision. Our findings contribute to the growing literature on effective formative assessment practices while providing practical insights into optimizing rubric implementation for enhanced student writing performance.

Rubrics and assessment

Instructional feedback has been shown to have a profound impact on student learning and performance (Hattie & Clarke, 2018; Wisniewski et al., 2020). Lipnevich and Smith (2022) defined feedback as “any information related to a performance that learners can use to improve their performance or learning. Feedback might come from any source, including teachers, peers, or the task itself. It may include information on where the learner is, where the learner is going, or what steps should be taken and strategies employed to improve, p. 76”. This conception of feedback maps directly onto Hattie and Timperley’s (2007) influential framework, which organises effective feedback around three core questions: “Where am I going?” concerning the goals and criteria for success; “How am I going?” providing information about current performance relative to those goals; and “Where to next?” offering guidance on how to close the gap between current and desired performance. Rubrics fit this definition as a rich form of feedback as it clearly shows individuals not only where they currently are, but also the level of performance that is expected from them while also providing guidance on how to reach those desired levels. Rubrics also enable students to generate self-feedback, which is a crucial element in their engagement with instructional feedback and the revision process (Nicol, 2021).

For instance, Andrade et al. (2008) provides empirical evidence of how rubrics can facilitate self-assessment among elementary students. In this quasi-experimental study, students in the treatment condition first analyzed model papers to develop writing criteria and then used rubrics to self-assess their initial drafts. Those who engaged in this process outperformed their peers in the control group in overall writing quality. Similarly, Andrade et al. (2010) extended this line of inquiry to a middle school context and once again participants in the treatment group showed significant improvements in writing quality across all rubric dimensions. Together, these studies show that rubrics are not merely evaluative tools, but instructional supports that activate self-regulatory processes fundamental to self-assessment and student learning. The effectiveness of rubrics in assessment practices can be increased by actively engaging students with this tool. For instance, having undergraduate students use a rubric when taking part in peer-feedback practices has been found to improve their

own performances (Greenberg, 2015). Moreover, the co-creation of rubrics has also been linked to positive outcomes, as co-creating rubrics allows students to work with their teachers in developing assessment criteria, fostering a deeper understanding of assessment expectations. This collaborative process aims to strengthen students' self-regulated learning skills and may lead to improved academic performance (Cockett & Jackson, 2018).

Rubric before task

Beyond their role as feedback instruments, rubrics also align closely with the framework of self-regulated learning (SRL), which emphasizes students' active role in planning, monitoring, and evaluating their own learning (Zimmerman, 2002). When presented before a task, rubrics can support the *forethought* phase of SRL by helping students set goals, plan strategically, and allocate cognitive resources. When used after a draft, rubrics can enhance the *self-reflection* phase, allowing students to evaluate their performance and identify areas for improvement. Thus, the timing of rubric presentation may differentially support distinct SRL processes (planning in the pre-task phase and reflection and revision in the post-task phase).

Even when students do not receive additional opportunities to engage with rubrics, such as in peer-assessments or in the co-creation of those tools, having access to this tool can lead to positive outcomes. Jonsson (2014) presented rubrics to higher education students before they began their tasks to clarify assessment expectations and guide them during planning and assessment processes. Students reported finding the rubrics useful and actively engaged with the criteria while working, which positively impacted their performance. Similarly, Greenberg (2015) found that participants who had access to a rubric before they started their assignments performed better than the ones who did not.

Rubrics can support various dimensions of student writing, including both reflective and academic writing, when introduced before or during the writing process. For instance, Cheng and Chan (2019) examined the use of scoring rubrics to enhance reflective writing among undergraduates and found that students who received rubrics, whether with or without detailed explanation, produced significantly higher levels of reflection than those in the control group, which did not have any access to a rubric. Although no difference emerged between the two treatment conditions, student interviews revealed that a lack of familiarity with rubrics limited their effectiveness, highlighting the importance of receiving explicit instruction on rubric-use. Similarly, Sundeen (2014) investigated the effects of rubric presentation on high school students' argumentative writing and found that both explicitly teaching the rubric and simply handing it out to students led to better performance than withholding it entirely. Together, these studies suggest that rubrics can improve writing quality by clarifying expectations and guiding students' task planning. However, they also argue for the importance of thoughtful instructional support to help students use rubrics effectively, especially when they lack prior experience with such tools.

Rubric after task - rubric as feedback

Research has demonstrated the effectiveness of rubrics as assessment tools, particularly when provided to students after they have produced initial drafts (Lipnevich et al., 2014, 2023). To investigate the optimal use of rubrics as feedback, Wollenschläger et al. (2016) conducted an experimental study examining the specific type of information rubrics should include to positively impact student performance. In this study, 120 grade 8 German students were randomly assigned to one of three feedback conditions. The Transparency Information condition provided students with a rubric outlining achievement levels for the tasks, allowing them to reflect on “Where am I going?” However, this condition lacked details on current performance or ways to improve. The Individual Performance Information condition gave students a rubric with an additional grade (ranging from 1 to 5) that compared their performance to peers, addressing “How am I going?” but without guidance for improvement. Finally, the Individual Performance Improvement Information condition presented students with a rubric, grade, and specific improvement cues that highlighted achieved and unmet rubric levels, answering “Where to next?” by helping students understand the steps needed to enhance their performance. The results of the study indicated that rubric feedback with targeted improvement information significantly improved students’ self-evaluation accuracy and led to better performance in subsequent scientific reasoning tasks. These findings suggest that students require additional information beyond the rubric itself to effectively engage in self-assessment practices.

Contrary to the findings of Wollenschläger et al. (2016), unmarked rubrics have been found to significantly enhance students’ performances (Lipnevich et al., 2014, 2023). Lipnevich et al. (2014) explored the impact of rubrics and exemplars on college students’ writing performances. In this experimental study, participants were asked to revise a written research proposal after receiving one of three instructional tools: a detailed rubric, exemplars of strong and weak proposals, or a combination of both. This study found that even though students in all conditions demonstrated improvements in assignments, it was the rubric-only group that showed the greatest improvement. Lipnevich et al. (2023) extended this investigation to a high school context, involving students from grades 9 and 10. Students wrote two essays and revised them based on the same feedback conditions as the earlier study. Consistent with the earlier findings, the rubric condition led to the most significant performance improvements, followed by the exemplars and the combined rubric-exemplar condition. After providing a training session on how to use these tools for the purposes of self-feedback generation, all three experimental conditions had a similar improvement in their performance. The researchers concluded that when it comes to rubrics, by providing clear criteria for performance they facilitate deep student engagement with the revision process. Moreover, the study revealed that rubrics help students generate self-feedback more effectively, whereas exemplars may require additional scaffolding for optimal use. Similarly, in the context of L2, Burnell et al. (2023) also found that both rubrics and exemplars demonstrated significant positive effects on performance and learning, with rubrics showing a tendency to be slightly more effective, especially in transferring skills to new tasks.

Similarly, in the context of English as a foreign language, Kim (2019) found a positive effect of rubric-referenced self-assessment accompanied by training high-school students on rubric-use. Participants had access to the rubric and training after the completion of an initial draft and both the quality and the length of their writing assignments were improved.

Moreover, qualitative data collected through questionnaires, interviews, and self-assessment diaries indicated that students perceived the training as effective, enhancing their understanding of writing criteria and fostering a more objective self-evaluation process.

Initial evidence to differential effects of rubrics depending on the timing it is presented to students has been found (Zhang et al., 2019). These authors explored the impact of the timing of rubric use on students' learning outcomes, metacognitive awareness, cognitive load, and attitudes in a flipped learning environment using a sample of 83 undergraduate students in China. In this quasi-experimental study, the experimental group received rubrics before starting their tasks, while the control group was given rubrics only after completing an initial draft. The results showed that only the control group made statistically significant improvements from Stage-I to Stage-II submissions, suggesting that receiving rubrics after an initial attempt may enhance students' ability to reflect and revise effectively. However, students in the experimental group reported higher metacognitive awareness and lower cognitive load, indicating that early rubric access supported planning, monitoring, and reduced mental demands. Both groups expressed generally positive attitudes toward rubric use, though some students noted that early access to rubrics might restrict creativity. These findings provide important preliminary evidence regarding the timing of rubric implementation and its varying impact on different dimensions of learning. As educators continue to explore ways to optimize the use of rubrics, gaining clarity on the most effective timing for their implementation will be essential. Whether rubrics are most beneficial when provided before, after, or at multiple stages of a task remains an open question that warrants deeper investigation. We addressed this question in our study.

Rubrics and cognitive load

Cognitive Load Theory (CLT) is based on the premise that working memory has a limited capacity, and effective learning occurs when cognitive load is managed appropriately. Contemporary CLT primarily distinguishes between two types of cognitive load: (1) Intrinsic Cognitive Load: which refers to the inherent difficulty associated with a specific task and varies depending on the complexity of the material and the learner's prior knowledge; (2) Extraneous Cognitive Load: which is the load affected by the way information is presented to learners, that is, it is dependent on the quality of the design of the instruction. (Chandler & Sweller, 1991; Kirschner, 2002; Paas et al., 2010; Sweller et al., 2011).

In a recent review of CLT, Sweller (2023) argues that germane cognitive load should no longer be treated as a separate type of load, but rather as the thinking efforts one exerts when trying to understand and organize new information. From this perspective, rubrics have the potential to significantly reduce cognitive load for students in several ways. First, they can reduce extraneous cognitive load by providing explicit criteria that clarify task expectations. This helps students to focus their attention on the essential elements of the assignment rather than expanding mental effort trying to decipher ambiguous requirements. Second, rubrics can lessen intrinsic cognitive load by breaking down complex tasks into more manageable parts. This is particularly beneficial for assignments that require multiple steps or specific formats, such as writing tasks. By structuring the work into discrete, measurable criteria, rubrics make the overall task feel less overwhelming for students. Finally, rubrics can support productive cognitive effort by promoting schema development. When

students can clearly recognize the essential elements of quality work through the rubric criteria, they are better able to apply that understanding to future tasks. This schema building process enhances students' capacity to self-regulate their performance and transfer learning to new contexts.

The mechanisms by which rubrics improve performance in writing tasks have been further explored in recent research. Krebs et al. (2022) investigated how rubrics enhance self-assessment accuracy and reduce cognitive load. Using a sample of high school students, the authors found that the rubric group showed significantly higher self-assessment accuracy and reported lower rates of difficulty when performing self-assessments compared to the control group, indicating that rubrics reduced cognitive load. Similarly, Zhang et al. (2019) reported a positive impact of rubrics on cognitive load. In their study, the experimental group, which had access to the rubric before starting the task, reported lower cognitive load, suggesting that engaging with rubrics helped students clarify the task expectations, which simplified their learning process.

Despite the increased interest in research concerning rubrics in the field of assessment, it is still unclear whether there is an optimal moment to present them to students. On one hand, having access to the rubrics before starting a task can make assessment criteria transparent (Brookhart, 2018) and potentially decrease cognitive load (Zhang et al., 2019). Providing students with the rubric upfront allows them to clearly understand the expectations and structure their work accordingly from the beginning. On the other hand, when students receive the rubric after they have completed an initial draft of their assignment, they are able to use that rubric as a comparator and generate self-feedback that can lead to enhanced learning and performance (Lipnevich et al., 2014, 2023). This post-task approach may encourage deeper engagement and reflection while reducing cognitive load.

The empirical support for the overall benefits of rubric use in assessment practices is compelling. Nevertheless, despite the strong evidence in favor of using rubrics in classroom assessment, there remains a lack of consensus concerning the ideal timing for presenting them to students. This gap in the literature underscores the need for further investigation to determine whether rubrics are most effective as pre-task guides or post-task feedback tools. Addressing this unresolved issue was the primary motivation for conducting the present study.

The current study

In this experimental study we examine the impact of rubric access on students' writing performance, cognitive load, and revision processes across two grade levels. By manipulating the timing of rubric access (i.e., before, after, both before and after, or not at all) we aim to determine how exposure to rubrics influences students' writing and cognitive processing. Given that rubrics serve as instructional tools that clarify expectations, their timing may play a crucial role in shaping students' engagement with feedback and revisions. This study extends prior research by investigating the relative effects of rubrics on writing outcomes at different stages of the writing process as well as students' cognitive load.

The following RQs guided this study:

1. Are there differences in writing performance across four conditions: no rubric, rubric before the initial draft, rubric after the initial draft, and rubric both before and after?
2. Are there differences in performance among grade levels?
3. Does cognitive load (perceived difficulty and effort) vary depending on the condition?

Several directional hypotheses could be formulated:

H1: The conditions with Rubric Before and Rubric Before and After will perform significantly better than conditions with Rubric After and No Rubric for the Draft (T1).

H2: The conditions with Rubric After and Rubric Before and After will perform significantly better than conditions with Rubric Before and No Rubric for the Revision (T2).

H3: The effects will be larger for the 11th grade than for the 6th grade, with 11th graders showing greater improvement.

Since there is limited research on the effect of our manipulation on the cognitive load of the students, our study has an exploratory nature regarding this objective.

Method

Participants

Participants were 271 students from a private school from the Northeast of the United States, with 141 (52.03%) enrolled in the 11th grade and 130 (47.97%) enrolled in the 6th grade. Our sample had a well-balanced gender distribution: 134 (49.44%) of the students self-identified as men and 137 (50.55%) self-identified as women. The mean age of 6th grade students was $M = 11.29$ ($SD = 0.47$), and for 11th grade students $M = 15.97$ ($SD = 0.38$). Of the total sample, 143 (52.77%) were White, 54 (19.92%) were Asian, 29 (10.70%) were Latino/a, 15 (5.54%) were Black, 20 (7.38%) were from other ethnicities, and 10 (3.33%) chose not to answer the question. Most participants ($n = 221$, 81.55%) were native speakers of English, whereas $n = 50$ (18.45%) had other languages as their first language.

Before conducting our study, we made an a priori estimate of the sample size required to achieve optimal statistical power. Based on the most restrictive analytical condition, we calculated the sample size needed to achieve a statistical power of 80% in a Repeated Measures Analysis of Variance with Between and Within interactions, with an estimated effect size of partial η^2 of 0.02, and assuming a Type I error rate of $\alpha = 0.05$, and a correlation between the two repeated measures of $r = .5$. The resulting minimum sample size estimate was $n = 183$. This power analysis was conducted using the *pwrrs* package (Bulum, 2023).

Procedure

This study was pre-registered following the guidelines of the Open Science Framework (OSF) to ensure research transparency. A project was created in the OSF repository on September 22, 2023 (https://osf.io/294p6/?view_only=8d6fabde4df84984934621f008ffd6b0), which served as the repository for all materials related to this work. On September 26, 2023, the study was pre-registered on the same platform (https://osf.io/23bpf/?view_only=4fb50afe760e4eadbb50c28baf73918b) and linked to the existing project. Ethical approval for the

study was obtained from the institution, under protocol 2022-0620-XX. Written consent for participation was secured from parents and students in accordance with ethical guidelines.

At the start of the study, all students were introduced to the assignment and the assessment criteria outlined in the rubric, with the exception of performance levels. Students were then instructed to write an initial essay draft and were randomly assigned to one of four experimental conditions (See Fig. 1).

1. *No Rubric* (NoR): Students did not have access to the rubric at any point during the study. This condition functioned as the control group.
2. *Rubric Before* (RB): Students were given access to the rubric before beginning their initial drafts.
3. *Rubric After* (RA): Students received access to the rubric after completing their initial drafts but before starting their revisions.
4. *Rubric Before and After* (RB&A): Students had access to the rubric both before starting their initial drafts and again after completing their drafts, just prior to revising.

In this study, random assignment was conducted for all students in grades 6 and 11, ensuring a balanced and unbiased distribution of students across conditions. This approach allowed for a robust examination of how grading and feedback practices vary across different classroom contexts. Importantly, data were not nested within classes, as random assignment ensured that students were distributed independently of their class groupings. Additionally, classes were taught by the same teacher in each grade, which helped control for teacher-specific effects and provided a clearer picture of how grading and feedback practices are influenced by student and teacher characteristics, as well as the interaction between them. The use of random assignment, combined with the consistent teaching structure within grades, strengthens the study's ability to draw causal inferences.

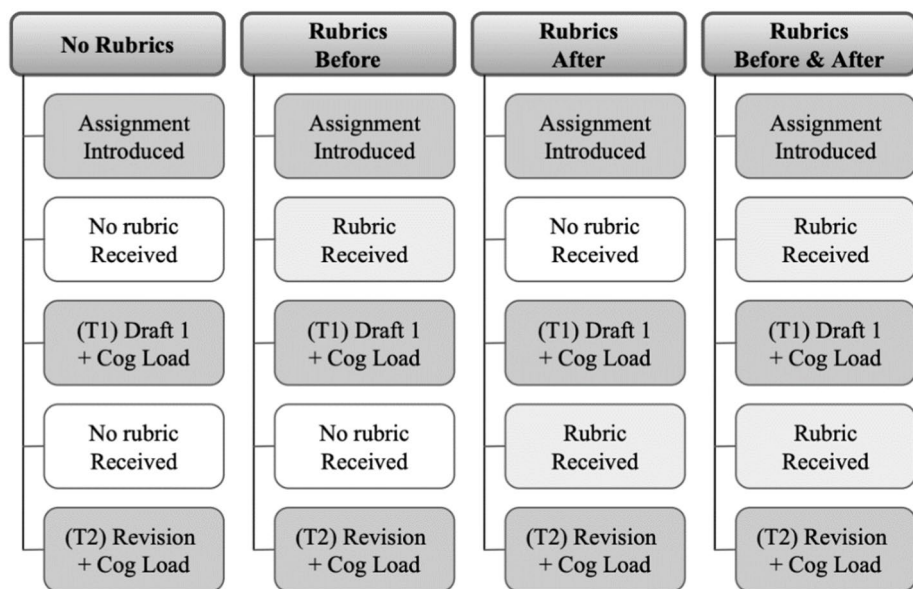


Fig. 1 The sequence of actions for each experimental condition

After completing their initial drafts, all students responded to two cognitive load questions. They then revised their drafts according to the rubric access specified by their assigned condition. Upon submitting their revised drafts, participants were asked to report their cognitive load again. All participants received the maximum possible score on the Grammar, Mechanics, and Formatting dimension of the rubric, as this criterion was not the focus of the intervention; accordingly, this dimension was excluded from the main analyses. A detailed overview of the experimental procedure for each condition is provided in Fig. 1.

The task

The task was integrated into regular instructional activities, and all students in grades 6 and 11 who were present on the day of the lesson participated. Sixth and eleventh-grade students were tasked with writing persuasive essays tailored to an appropriate level of difficulty.

For the 6th graders, the prompt was: “Should homework be banned in elementary schools?” Students were asked to take a clear stance on whether homework is beneficial or harmful for young learners and to provide well-reasoned arguments to support their views. This task encouraged them to think critically about the role of homework in their education and to articulate their thoughts clearly.

For the 11th graders, the prompt was: “Should high schools provide more vocational training options?” Here, students were tasked with persuading their audience about the benefits of increasing vocational training in high schools and how it could prepare students for successful careers. This prompt required students to consider the diverse needs of future workers and how vocational training might better align education with employment opportunities.

Both essay prompts were designed not only to assess students’ writing abilities but also to challenge them to think deeply about the relevance and impact of their education on their future paths.

Table 1 shows the 2×2 repeated measures factorial experimental design where the use of the rubric before the essay writing task is manipulated on one hand, and the timing of the rubric presentation on the other. As depicted in Table 1, the counterbalancing of this factorial design results in four groups labeled: *No rubrics*, *Rubrics before*, *Rubrics after*, and *Rubrics before and after*. In each of the two treatment phases, subjects in the groups either receive (marked as R) or do not receive (marked as -) the rubric. In each phase, after the treatment, all subjects complete a writing task/revision (T) and then report on the cognitive load (CL) experienced during that phase. This factorial design was implemented with both 6th and 11th graders.

Table 1 Experimental design

Group	Time 1		Time 2	
	Treatment	Measures	Treatment	Measures
NoR	–	Draft CL	–	Revision CL
RB	R ₁	Draft CL	–	Revision CL
RA	–	Draft CL	R ₂	Revision CL
RB&A	R ₁	Draft CL	R ₂	Revision CL

Note. NoR: No Rubric; RB: Rubric Before; RA: Rubric After; and RB&A: Rubric Before and After. R: Use of rubric in phase before working on a draft, number in sub-index indicates phase; -: no treatment. CL: cognitive load (Difficulty and Effort)

Measures

Writing performance scores

Both the draft and revision were scored on a 1 to 100 scale by two raters using a rubric. Throughout this paper, scores from the initial writing task are referred to as Draft scores (T1), and scores from the revised writing task are referred to as Revision scores (T2), consistent with the labels used in all tables and figures. The rubric included five performance criteria with assigned point values, and the total score was used for analysis. All essays were scored by two raters, in line with the school's grading policy. Schools have been using the rubric for some time, with teachers participating in regular calibration sessions. To assess inter-rater reliability, Krippendorff's alpha was calculated, yielding values of 0.89 for 6th grade and 0.91 for 11th grade, both exceeding the commonly accepted 0.80 threshold for strong agreement (Krippendorff, 2004). The rubric's criteria and performance levels are presented in the Appendix.

Cognitive load

Cognitive Load was measured using single-item scales to assess two dimensions: students' perceived difficulty and the amount of effort exerted to complete the tasks. These two dimensions are labelled Difficulty and Effort throughout all tables and figures in this paper. After completing the initial draft, participants answered both *"How difficult was it to produce a draft of this essay?"* and *"How much effort did you invest into producing this draft?"* using a 7-point Likert scale. After revisions were completed, again using a 7-point Likert scale, students answered *"How difficult was it to revise this essay?"* and *"How much effort did you invest into revising this draft?"*. Single-item measures of cognitive load are widely used in educational research and have demonstrated adequate psychometric properties. Numerous studies have employed single-item scales to assess perceived difficulty and effort (e.g., Ayres, 2006; DeLeeuw & Mayer, 2008; Schmeck et al., 2015). Schuessler et al. (2024) specifically investigated the construct validity of single-item cognitive load ratings and found evidence that these items capture the intended constructs meaningfully, with acceptable levels of convergent validity. Moreover, a meta-analysis by Hodson (2016) found that single-item and multiple-item scales measuring cognitive load did not differ in measurement sensitivity, suggesting that single-item measures are not systematically inferior to longer scales for detecting differences across conditions. In the present study, participants completed these measures at two time points. The two items used here, namely perceived difficulty and effort, are labeled as Difficulty and Effort in all tables and figures, corresponding directly to the administered questions.

Data analyses

The analyses in this study were conducted using JASP (JASP Team, 2024) and R software (R Core Team, 2020).

The main between-groups variable was students' access to a rubric, with four conditions: No Rubric (NoR), Rubric Before (RB), Rubric After (RA), and Rubric Before and After (RB&A). The second between-group variable was students' educational level, with two

grade levels (6th graders and 11th graders). The within-subject variable was Time (T1 vs. T2), representing the transition from initial draft to revision (see Table 2).

To test the hypotheses more directly, we examined theoretically relevant contrasts between conditions, rather than relying solely on omnibus tests. In particular, for T1 performance, we focused on comparisons between conditions that differed in rubric availability at the drafting stage (e.g., NoR vs. RA, and RB vs. RB&A). For T2 performance, we examined contrasts comparing each rubric condition with the control condition (NoR), as well as comparisons between rubric timing conditions (e.g., RA vs. RB). These contrasts reflect the theoretical structure of the experimental manipulation and allow for a more precise evaluation of the hypotheses.

Based on our first research question (RQ1), we formulated two hypotheses. In H1, we tested whether the *Rubric Timing* condition affected students' performance on the Draft assignment (T1). In H2, we examined the effect of *Rubric Timing* on students' performance on the Revised assignment (T2).

In RQ2, we explored whether Grade level (6th vs. 11th) moderated the effects proposed in H1 and H2. Specifically, H3 predicts that the effects of Rubric Timing will be larger for 11th-grade students than for 6th-grade students, with older students showing greater improvement.

RQ3 addressed whether students' perceived cognitive load (measured as perceived difficulty and effort) varied across conditions. Given the limited prior research on how Rubric Timing influences cognitive load, this research question is exploratory in nature.

Our proposed analytical plan was as follows:

To test H1, we examined differences in Draft (T1) scores across the four Rubric Timing conditions:

- *No Rubric (NoR)*.
- *Rubric Before (RB)*.
- *Rubric After (RA)*.
- *Rubric Before & After (RB&A)*.

Table 2 Summary of research questions, hypotheses, variables, contrasts, and statistical models

RQ	Hypothesis	Outcome	Predictors	Key contrasts	Model
RQ1	H1	Draft (T1)	Rubric condition	NoR vs. RA; RB vs. RB&A	Binomial logistic regression
RQ1	H2	Revision (T2)	Rubric condition	RB vs. NoR; RA vs. RB; RA vs. RB&A	Binomial logistic regression
RQ2	H3	Draft (T1), Revision (T2)	Grade level × Rubric condition	Interaction contrasts by grade	Binomial logistic regression
RQ2	–	T2 (adj. T1)	Grade level × Rubric + T1	Adjusted group comparisons	Binomial logistic regression
RQ3	–	Difficulty	Rubric condition, Grade level	Differences across conditions	Ordinal regression
RQ3	–	Effort	Rubric condition, Grade level	Differences across conditions	Ordinal regression

To test H2, we used the same comparison across conditions for Revision (T2) scores. To test H3, we added Grade level (6th vs. 11th) as a predictor to investigate its interaction with Rubric Timing in influencing Draft (T1) and Revision (T2) performance. To evaluate whether T2 performance was influenced by T1 performance, we conducted an additional model including Draft T1 performance as a covariate to account for baseline differences. For RQ3, we repeated the above analyses using cognitive load measures (effort and perceived difficulty) as outcome variables.

Our original analytic plan specified 2 (Grade level) $\times$ 4 (Rubric Timing) ANOVAs on Draft (T1) and Revision (T2) scores. However, given the bounded nature of the outcome variable (scores ranging from 0 to 100), we considered that models assuming normally distributed residuals might be suboptimal. Diagnostic analyses (see Appendix A) indicated that deviations from normality were generally modest, but there were indications of heterogeneity of variances, particularly in models adjusting for baseline performance (T1). In addition, the distribution of scores showed some clustering at higher values, suggesting potential ceiling effects. Accordingly, we opted to use a binomial logistic regression approach, which models performance relative to the maximum possible score and is more appropriate for bounded outcomes. This approach is also more robust to heteroscedasticity. In this analysis, both T1 and T2 scores were transformed into success/failure outcomes.

For the cognitive load measures, which were ordinal in nature, we applied ordinal regression models to better fit the data's distributional properties, in line with recommendations from Bürkner and Vuorre (2019) and Liddell and Kruschke (2018).

The significance level (α) for all statistical tests was set at 0.05. These comparisons provided deeper insights into how each level of the counterbalanced treatment influenced the observed outcomes.

Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and the study follows JARS (Applebaum et al., 2018). The raw data, as well as the original and revised version of the analytical plan, can be found at (https://osf.io/294p6/?view_only=8d6fabde4df84984934621f008ffd6b0). Data were analyzed using R, version 4.0.0 (R Core Team, 2020) and the package *ggplot*, version 3.2.1 (Wickham, 2016). This study's design and its analysis were pre-registered prospectively (see above for details).

Results

Table 3 presents the descriptive statistics for the dependent variables (T1 Draft, T2 Revision, Difficulty and Effort) as a function of the experimental condition. No outliers (defined as values exceeding ± 3 standard deviations) were identified for any of the dependent variables.

Table 4 displays the correlations among the dependent variables across the two measurement points. Table presenting correlations by condition is presented in the Supplementary.

To test H1, we fitted a binomial logistic regression model predicting success in Draft T1, including Grade Level (6th vs. 11th) and Rubric Timing (NoR, RB, RA, RB&A) as predictors. No significant effects were observed. Consistent with the theoretically relevant

Table 3 Descriptive statistics by experimental condition

Condition	Time	VD	Valid	Missing	M	SD	Minimum	Maximum
NoR								
	T1	Draft	66	0	74.21	8.15	55.00	93.00
		Difficulty	66	0	5.62	1.60	1.00	7.00
		Effort	66	0	6.24	0.88	4.00	7.00
	T2	Revision	66	0	74.21	8.72	55.00	96.00
		Difficulty	67	1	4.04	2.36	1.00	7.00
		Effort	66	0	4.17	2.49	1.00	7.00
RB								
	T1	Draft	69	0	74.39	8.23	58.00	91.00
		Difficulty	69	0	5.45	1.69	1.00	7.00
		Effort	69	0	6.39	0.83	4.00	7.00
	T2	Revision	69	0	76.71	8.44	55.00	96.00
		Difficulty	69	0	4.07	2.40	1.00	7.00
		Effort	69	0	4.54	2.43	1.00	7.00
RA								
	T1	Draft	68	0	73.79	8.59	55.00	93.00
		Difficulty	68	0	5.46	1.48	1.00	7.00
		Effort	68	0	6.16	1.10	1.00	7.00
	T2	Revision	68	0	81.50	8.48	55.00	94.00
		Difficulty	68	0	4.10	2.28	1.00	7.00
		Effort	68	0	3.76	2.45	1.00	7.00
RB&A								
	T1	Draft	68	0	74.91	8.24	55.00	93.00
		Difficulty	68	0	5.09	1.82	1.00	7.00
		Effort	68	0	6.29	1.11	1.00	7.00
	T2	Revision	68	0	78.32	7.45	59.00	96.00
		Difficulty	68	0	4.41	2.33	1.00	7.00
		Effort	67	1	4.04	2.36	1.00	7.00

Note. NoR: Group No Rubric; RB: Group Rubric Before; RA: Group Rubric After; and RB&A: Group Rubric Before and After. T1: Time 1, T2: Time 2

Table 4 Pearson correlations for the dependent variables as a function of the time of measurement

Variable	Draft T1	Revision T2	Difficulty T1	Effort T1	Difficulty T2	Effort T2
Draft T1	–					
Revision T2	0.76***	–				
Difficulty T1	0.10	–0.01	–			
Effort T1	0.05	0.05	0.29***	–		
Difficulty T2	–0.03	–0.11	0.10	–0.06	–	
Effort T2	0.00	0.11	–0.25***	0.14*	–0.79***	–

Note. T1: Time 1, T2: Time 2. * $p < .05$, ** $p < .01$, *** $p < .001$

contrasts specified a priori, no differences were found between conditions that did not differ in rubric availability at T1 (i.e., NoR vs. RA, and RB vs. RB&A). For the cognitive load variables (Difficulty and Effort), a similar model was fitted using ordinal regression, as part of the exploratory analyses for RQ3. Only for Difficulty was the main effect of Grade level

significant, $\chi^2 = 8.70$, $p = .003$ showing a higher level of difficulty perceived by 6th grade students compared to 11th grade students ($OR = 1.35$, $p = .003$).

To test H2, we performed the same analysis for the variables evaluated at Time 2. The binomial logistic regression model for success in Revision T2 showed a main effect of Grade level, $\chi^2 = 27.94$, $p < .001$, a main effect of Rubric Timing, $\chi^2 = 111.44$, $p < .001$, and the Grade $\times$ Rubric Timing interaction, $\chi^2 = 15.30$, $p = .002$.

Among 11th-grade students, all rubric conditions outperformed the NoR condition (all p 's $< .001$). However, using the rubric after the task (RA) resulted in the greatest improvement (83.2% success rate for RA vs. 73.9% for NoR, $OR = 1.75$, $p < .001$). In addition, RA also showed improvements over RB ($OR = 1.38$, $p < .001$).

In the case of 6th-grade students, although the use of a rubric was associated with improved performance, the pattern of improvement was more limited. Specifically, RA showed a significant improvement compared to NoR (78.8% vs. 74.5%, $OR = 1.33$, $p < .001$), whereas neither RB nor RB&A differed significantly from NoR. These results are consistent with the planned contrasts, showing that the RA condition outperformed both the control condition (NoR) and other rubric timing conditions, particularly among 11th-grade students. Detailed descriptive statistics disaggregated by grade level are provided in Appendix B, Table B3.

The ordinal model for Difficulty T2 showed a main effect of Grade, $\chi^2 = 72.61$, $p < .001$, with 6th-grade students reporting greater perceived difficulty than 11th-grade students ($OR = 87.05$, $p < .001$). For Effort T2, the main effect of Grade was also significant, $\chi^2 = 81.22$, $p < .001$, reflecting the inverse pattern, with 11th-grade students reporting greater effort ($OR = 0.004$, $p < .001$). Full model estimates and pairwise contrasts are reported in Appendix B (Table B1).

Figure 2 illustrates the results of post hoc analyses, depicting the average scores obtained by students at both time points (Draft T1 and Revision T2). The horizontal axis represents time, lines correspond to grade level, and panels represent rubric conditions (NoR, RB, RA, and RB&A).

The binomial logistic regression model (adjusted for initial performance at Draft T1) on student success at Revision T2 showed the main effect of Grade, $\chi^2 = 23.52$, $p < .001$, the main effect of Rubric Timing, $\chi^2 = 127.45$, $p < .001$, and a significant interaction between Grade by Rubric Timing, $\chi^2 = 14.12$, $p = .003$, which suggests that there is a different pattern in the way the rubric affected performance in 6th and 11th grade students. Specifically, in the 11th grade students, those in the RA condition showed superior success in T2 Revision compared to the NoR condition ($OR = 0.57$, $p < .001$), as with RB ($OR = 1.41$, $p < .001$), and RB&A ($OR = 1.25$, $p = .003$). In contrast, in the 6th grade students, only the RA condition was shown to improve success over the control condition ($OR = 0.71$, $p < .001$). This pattern is consistent with the theoretically relevant contrasts, indicating that the effect of rubric timing differed across grade levels. Among 11th-grade students, all rubric conditions outperformed the control condition, with the strongest effect observed for the RA condition. In contrast, among 6th-grade students, only the RA condition showed improved performance relative to the control condition. Full model estimates and pairwise contrasts are reported in Appendix B, Table B2.

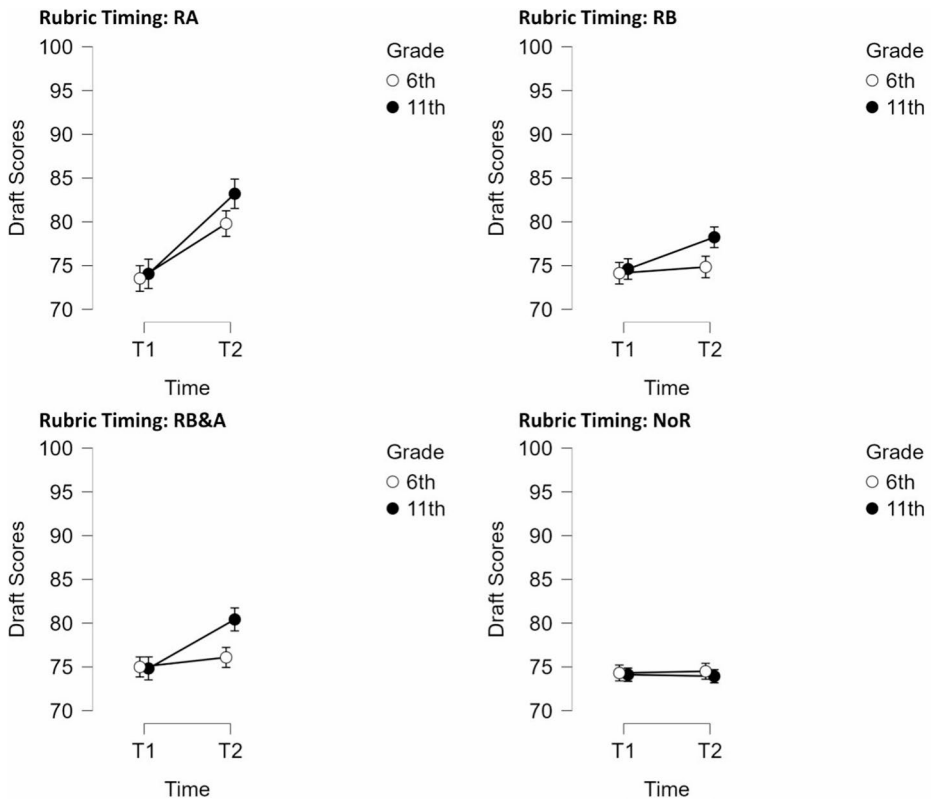


Fig. 2 Draft results as a function of time, type of rubric timing, and grade. *Note:* Error bars represent 95% CI

Discussion

The goal of this study was to explore how the timing of rubric presentation influences student performance and the underlying cognitive mechanisms that drive these effects. Specifically, we examined whether providing rubrics before starting a task, after completing an initial draft, or at both stages would lead to different outcomes in terms of final draft quality and student engagement with the feedback process. By addressing this gap in the literature, our study aimed to determine whether rubrics function more effectively as pre-task guides or post-task feedback tools. The findings demonstrate that the timing of rubric presentation significantly impacts how students engage with the writing process, shaping both their approach to the task and their ability to refine their work during revision. These results contribute to a deeper understanding of how rubrics can be optimized to enhance learning outcomes and inform best practices in formative assessment.

We found that overall, for both grade levels, presenting the rubric *after* the initial draft of the work is completed leads to significantly greater improvement in performance compared to presenting the rubric either *before* the task or both *before and after*. This finding suggests that rubrics may be more effective as post-task feedback tools, serving as a framework for self-assessment and revision. These results are consistent with findings from Lipnevich et

al. (2014, 2023; Zhang et al., 2019) which demonstrated the effectiveness of rubrics for self-feedback generation. By engaging with the rubric after completing a draft, students may better identify discrepancies between their work and the desired criteria, fostering deeper reflection and more targeted improvements. This highlights the potential of strategically timed rubric use to enhance learning and performance.

Hypotheses and performance differences across rubric conditions

In our pre-registration, we hypothesized that students in the *Rubric After* condition would outperform those in other conditions on their revised essays. This hypothesis was supported by the data, demonstrating the effectiveness of providing rubrics as a post-draft feedback tool. By engaging with the rubric after completing an initial draft, students likely utilized it as a reflective mechanism, enabling them to identify and address deficiencies in their work. Having access to the rubric after completing an initial draft enabled students to generate internal feedback through a process of comparison. That is, it allowed students to compare their current performance against a reference point (Nicol, 2021). This aligns with prior research suggesting that rubrics presented post-task encourage self-assessment and deeper reflection (Lipnevich et al., 2014, 2023). However, contrary to our expectations, we observed no significant differences in performance on the initial drafts among the *Rubric Before* and *Rubric Before and After* conditions. This finding challenges the notion that providing a rubric before writing directly enhances the quality of initial drafts, suggesting that the immediate availability of assessment criteria may not necessarily translate into improved draft performance. Cheng and Chan (2019) found that both experimental groups, which had access to rubrics (with and without explanations), scored higher in reflection writing quality compared to the control group. Our finding also contrasts with recent meta-analysis, which has indicated an overall moderate positive effect of rubrics on students' academic performance (Panadero et al., 2023). One possible explanation is that students might not have internalized or effectively utilized the rubric during the drafting phase, highlighting the need for explicit instruction or scaffolding on how to apply rubrics as pre-task guides. For rubrics to be effective, students must engage with them. For instance, Becker (2016) found that different levels of quality of engagement with rubrics led to different levels of performance improvement: students who either participated in co-creating a scoring rubric or used a scoring rubric to assess writing performance had significantly better overall writing quality than did students who simply saw the rubric or belonged to the control group.

Grade-level differences in improvement

We also hypothesized that 11th-grade students would demonstrate greater improvements than 6th-grade students, reflecting their more advanced writing skills and greater familiarity with feedback processes. This hypothesis was largely supported by the data, as 11th-grade students showed more substantial overall improvements between Draft T1 and Revision T2 compared to 6th-grade students. This finding is in line with the moderation analysis reported in Panadero et al. (2023), which suggested that rubrics were more beneficial to secondary students compared to learners from earlier educational levels. Interestingly, however, an exception emerged in the *Rubric After* condition, where both 6th and 11th-grade students exhibited equally large improvements. This finding underscores the particular utility of post-

draft rubric feedback across developmental stages, suggesting that even younger students, who may struggle with self-assessment or feedback application, can benefit significantly from targeted guidance provided after completing an initial draft. Support to the potential of rubrics as feedback for learners as young as first and second graders was also provided by Bradford et al. (2016). Their results showed that providing instruction on the skills outlined by the rubric, paired with student access to the rubric, improved student writing scores with a large effect size. The equal improvement in this condition may also point to the simplicity and clarity of rubric-based feedback, making it accessible to less experienced writers. This finding opens avenues for further investigation into how developmental factors interact with feedback timing to influence learning outcomes.

Cognitive load and effort across grades and conditions

Initial research on the impact of rubric on cognitive load has shown optimistic results with students reporting lower cognitive load when they had access to rubrics (Krebs et al., 2022; Zhang et al., 2019). However, comparisons of differences in cognitive load by students' educational levels when they have access to rubrics is lacking. Our study also revealed grade-level differences in cognitive load and effort. Sixth-grade students reported significantly higher difficulty with the writing task compared to 11th-grade students, which may reflect developmental differences in writing proficiency and experience. Notably, this perceived difficulty did not vary significantly across rubric conditions, suggesting that the timing of rubric presentation did not influence students' perceptions of task complexity. Conversely, 11th-grade students reported higher levels of effort, particularly in the *Rubric Before* condition. However, this increased effort did not translate into improved performance, indicating that effort alone is insufficient to drive better outcomes without effective guidance or scaffolding. This finding underscores the importance of aligning instructional tools with students' developmental needs and capabilities to maximize their potential impact. The discrepancy between effort and outcomes in the *Rubric Before* condition further highlights the need to explore how rubrics are introduced and integrated into the writing process to ensure their effective use.

Practical implications

Taken together, these findings have important implications for instructional design and formative assessment practices. The superior performance of the *Rubric After* condition suggests that rubrics are most effective when used as tools for feedback and revision rather than as pre-task guides. However, the lack of significant differences in draft performance across conditions points to the need for additional support when introducing rubrics upfront. Educators may consider pairing rubrics with explicit teaching on how to apply assessment criteria during the drafting process. Furthermore, the grade-specific findings highlight the necessity of tailoring feedback approaches to students' developmental levels. While younger students may struggle with writing tasks, offering clear guidance for improvement may facilitate their growth. Future research should explore how to optimize the integration of rubrics into various stages of the writing process and examine their long-term impact on learning outcomes.

Limitations and future directions

Although this study contributes valuable insights into the impact of rubric timing on student performance and feedback engagement, limitations must be acknowledged. First, the study focused on only two grade levels, 6th and 11th grades, which provided the opportunity to examine how age and writing experience may moderate the effects of rubric use. Our findings indicate that grade level plays a meaningful role in how students respond to rubric timing, an important contribution to the literature. However, we must be careful about generalizing these findings. Further research is needed to determine whether similar patterns hold across other age groups. Younger students or college-level learners may respond differently to the timing of rubric presentation, and future research should explore these additional populations.

Second, the study was conducted in a controlled experimental setting, which, while useful for isolating variables, may not fully replicate the complexities of real classroom environments. Factors such as teacher-student interactions, peer collaboration, and time constraints could influence the effectiveness of rubric use in authentic educational settings. Therefore, future research is needed to examine additional variables to enrich these findings.

Third, although we observed significant differences in performance improvement across rubric timing conditions, the mechanisms underlying these effects remain unclear. Employing additional qualitative measures, such as think-aloud protocols or interviews, could help clarify these underlying processes.

Fourth, the assignment focused specifically on persuasive essays, which may have influenced how students interacted with the rubrics and benefited from their timing. Persuasive writing requires a clear structure, argumentation, and evidence, which are elements that align well with rubric criteria and may have made the feedback more actionable. However, the impact of rubric timing could differ across other genres such as narrative, descriptive, or expository writing, which may emphasize different skills or cognitive demands. Future studies should examine whether the timing effects observed in this study generalize to other types of writing tasks to better understand how task characteristics interact with assessment tools like rubrics.

Despite these limitations, the study provides important contributions to the field of formative assessment and highlights the potential benefits of tailoring rubric use to the timing and developmental needs of learners. Future investigations should address these limitations to build a more comprehensive understanding of effective rubric implementation.

Conclusion

This study makes a significant contribution to the growing body of research on rubric use in assessment by providing comprehensive evidence that rubrics are most effective when used as a feedback tool *after* students have completed their drafts. To our knowledge, this is the first study to thoroughly investigate the impact of rubric timing in such a detailed manner, comparing conditions where rubrics are presented before, after, or both before and after the drafting process. Our findings suggest that rubrics provided as feedback (following the initial draft) are particularly beneficial in enhancing student performance on revisions. This timing facilitates a reflective process where students can better assess their work and iden-

tify areas for improvement, leading to more substantial gains in learning. Rubrics presented *before* the initial draft is written did not result in writing improvement.

The results underscore the importance of rubrics not as tools for setting expectations but as integral components of the feedback loop that can drive student progress. While rubrics before the task may provide clarity on expectations, our findings suggest that their true potential is realized when students use them as a guide for revising and improving their work. This insight offers valuable guidance for educators seeking to optimize the impact of rubrics in their teaching practices.

Moreover, the study highlights the critical role of timing in formative assessment. By demonstrating the effectiveness of rubrics as post-task feedback, we hope to encourage educators to rethink how and when they use rubrics to support student learning. This is the first study, to our knowledge, to comprehensively examine this aspect, and it provides a foundation for future research to explore the most effective strategies for rubric implementation. Ultimately, this research calls attention to the need to prioritize rubrics as feedback tools to foster deeper student engagement, reflection, and, most importantly, academic growth.

Appendix

Argumentative Essay Rubric (100 Points Total).

1. Thesis Statement and Focus (20 Points)

20–18 Points: The thesis is clear, specific, and strongly presents the writer's position on the issue. The essay remains tightly focused on the thesis throughout.

17–15 Points: The thesis is clear and presents a position, but it may lack some specificity or strength. The essay generally stays focused but may occasionally drift.

14–12 Points: The thesis is present but vague or overly broad. The essay sometimes loses focus or struggles to maintain a clear connection to the thesis.

11–0 Points: The thesis is unclear, missing, or does not present a position. The essay lacks focus and fails to address a central argument.

2. Argument Development and Evidence (30 Points)

30–27 Points: The argument is well-developed, with strong, relevant, and sufficient evidence to support the thesis. Counterarguments are addressed effectively, and the reasoning is logical and persuasive.

26–23 Points: The argument is developed with relevant evidence, but some points may lack depth or support. Counterarguments are acknowledged but not fully addressed.

22–18 Points: The argument is underdeveloped, with limited or weak evidence. Counterarguments are missing or poorly addressed, and the reasoning may be unclear or illogical.

17–0 Points: The argument is poorly developed or missing. Evidence is insufficient, irrelevant, or absent. Counterarguments are ignored, and the reasoning is flawed or nonexistent.

3. Organization and Structure (20 Points)

20–18 Points: The essay is well-organized, with a logical flow of ideas. Transitions between paragraphs are smooth, and the introduction, body, and conclusion are effectively structured.

17–15 Points: The essay is organized, but some transitions may be awkward or unclear. The structure is generally sound but could be more polished.

14–12 Points: The essay lacks clear organization, with disjointed or confusing transitions. The structure may feel incomplete or unbalanced.

11–0 Points: The essay is disorganized, with no clear structure or logical flow. Transitions are missing, and the essay is difficult to follow.

4. Language and Style (15 Points)

15–14 Points: The language is clear, precise, and appropriate for the audience. The tone is consistent and persuasive, and word choice enhances the argument.

13–12 Points: The language is clear but may lack precision or variety. The tone is generally consistent, but word choice may be repetitive or uninspired.

11–9 Points: The language is unclear or overly simplistic. The tone may be inconsistent, and word choice may detract from the argument.

8–0 Points: The language is confusing, inappropriate, or overly informal. The tone is inconsistent or inappropriate, and word choice significantly weakens the argument.

5. Grammar, Mechanics, and Formatting (15 Points)

15–14 Points: The essay is free of grammatical, spelling, and punctuation errors. Formatting (e.g., citations, paragraphing) is consistent and correct.

13–12 Points: The essay has minor errors in grammar, spelling, or punctuation that do not significantly detract from readability. Formatting is mostly consistent.

11–9 Points: The essay has frequent errors in grammar, spelling, or punctuation that occasionally hinder readability. Formatting may be inconsistent or incorrect.

8–0 Points: The essay has numerous errors in grammar, spelling, or punctuation that significantly hinder readability. Formatting is inconsistent or incorrect.

Total Points: 100.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11251-026-09821-3>.

Author contribution Anastasiya Lipnevich: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing - original draft; Writing - review & editing. Ligia Tomazin: Writing - original draft; Writing - review & editing. Samuel Parra León: Resources; Software; Supervision; Validation; Visualization; Data analysis. AI wasn't used during preparation of this manuscript.

Funding There was no funding to support this research. There is no conflict of interest to report.

Declarations

Competing interests The authors declare that they have no competing interests.

References

- Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education*, 5(3), 197–203. <https://doi.org/10.1187/cbe.06-06-0168>
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–19.

- Andrade, H. L. (2023). What is next for rubrics? A reflection on where we are and where to go from here. In J. E. Groccia, M. S. Alsudairi, & W. T. O'Brien (Eds.), *Improving learning through assessment rubrics: Student awareness of what and how they learn* (pp. 314–326). IGI Global. <https://doi.org/10.4018/978-1-6684-5988-5.ch016>
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement Issues and Practice*, 27(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2008.00118.x>
- Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles Policy & Practice*, 17(2), 199–214. <https://doi.org/10.1080/09695941003696172>
- Applebaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Atkinson, D., & Lim, S. L. (2013). Improving assessment processes in higher education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, 29(5). <https://doi.org/10.14742/ajet.526>
- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, 16(5), 389–400. <https://doi.org/10.1016/j.learninstruc.2006.09.001>
- Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing*, 29, 15–24. <https://doi.org/10.1016/j.asw.2016.04.001>
- Bradford, K., Newland, A., Rule, A., & Montgomery, S. (2016). Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. *Early Childhood Education Journal*, 44, 463–472. <https://doi.org/10.1007/s10643-015-0727-0>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2018.00022.3>
- Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368. <https://doi.org/10.1080/00131911.2014.929565>
- Bulum, B. (2023). *pwrrs: Statistical power and sample size calculation tools* (R package version 0.3.1). <https://CRAN.R-project.org/package=pwrrs>
- Bürkner, P. C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- Burnell, K., Pratt, K., Berg, D. A. G., & Smith, J. K. (2023). The influence of three approaches to feedback on L2 writing task improvement and subsequent learning. *Studies in Educational Evaluation*, 78, 101291. <https://doi.org/10.1016/j.stueduc.2023.101291>
- Camargo Salamanca, S. L. C., Parra-Martínez, A., Chang, A., Maeda, Y., & Traynor, A. (2024). The effect of scoring rubrics use on self-efficacy and self-regulation. *Educational Psychology Review*, 36(3), 70. <https://doi.org/10.1007/s10648-024-09906-w>
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. <https://www.jstor.org/stable/3233596>
- Cheng, M., & Chan, C. (2019). An experimental test: Using rubrics for reflective writing to develop reflection. *Studies in Educational Evaluation*, 61, 176–182. <https://doi.org/10.1016/J.STUEDUC.2019.04.001>
- Chowdhury, F. (2018). Application of rubrics in the classroom: A vital tool for improvement in assessment, feedback and learning. *International Education Studies*, 12(1), 61. <https://doi.org/10.5539/ies.v12n1p61>
- Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. *Nurse Education Today*, 69, 8–13. <https://doi.org/10.1016/j.nedt.2018.06.022>
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223–234. <https://doi.org/10.1037/0022-0663.100.1.223>
- English, N., Robertson, P., Gillis, S., & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, 113, 101964. <https://doi.org/10.1016/j.ijer.2022.101964>
- Ghaffar, M., Khairallah, M., & Salloum, S. (2020). Co-constructed rubrics and assessment for learning: The impact on middle school students' attitudes and writing skills. *Assessing Writing*, 45, 100468. <https://doi.org/10.1016/j.asw.2020.100468>

- Greenberg, K. P. (2015). Rubric use in formative assessment: A detailed behavioral rubric helps students improve their scientific writing skills. *Teaching of Psychology*, 42(3), 211–217. <https://doi.org/10.1177/0098628315587618>
- Hattie, J., & Clarke, S. (2018). *Visible Learning: Feedback*. Routledge. <https://doi.org/10.4324/9780429485480>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hodson, J. L. (2016). *Measuring cognitive load: A meta-analysis of load measurement sensitivity* (Publication No. 10157387). [Doctoral dissertation, The University of Utah]. ProQuest Dissertations & Theses Global.
- JASP Team (2024). *JASP* (Version 0.18) [Computer software]. <https://jasp-stats.org>
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840–852. <https://doi.org/10.1080/02602938.2013.875117>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kim, J. (2019). Effects of rubric-referenced self-assessment training on Korean high school students' english writing. *English Teaching*, 74(3), 79–111. <https://eric.ed.gov/?id=EJ1288090>
- Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12(1), 1–10. [https://doi.org/10.1016/S0959-4752\(01\)00014-7](https://doi.org/10.1016/S0959-4752(01)00014-7)
- Krebs, R., Rothstein, B., & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning*, 17(2), 627–650. <https://doi.org/10.1007/s11409-022-09302-1>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.
- Lerdpornkulrat, T., Poondej, C., Koul, R., Khiawrod, G., & Prasertsirikul, P. (2019). The positive effect of intrinsic feedback on motivational engagement and self-efficacy in information literacy. *Journal of Psychoeducational Assessment*, 37(4), 421–434. <https://doi.org/10.1177/0734282917747423>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Lipnevich, A. A., & Smith, J. K. (2022). Student-feedback interaction model: Revised. *Studies in Educational Evaluation*, 75, 101208. <https://doi.org/10.1016/j.stueduc.2022.101208>
- Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, 42(4), 539–559. <https://doi.org/10.1007/s11251-013-9299-9>
- Lipnevich, A. A., Panadero, E., & Calistro, T. (2023). Unraveling the effects of rubrics and exemplars on student writing performance. *Journal of Experimental Psychology: Applied*, 29(1), 136–148. <https://doi.org/10.1037/xap0000434>
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756–778. <https://doi.org/10.1080/02602938.2020.1823314>
- Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, 22(2), 115–121. <https://doi.org/10.1007/s10648-010-9133-8>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles Policy & Practice*, 21(2), 133–148. <https://doi.org/10.1080/0969594X.2013.877872>
- Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, 35(4), 113. <https://doi.org/10.1007/s10648-023-09823-4>
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, 7(1), 9. <https://doi.org/10.1186/s40594-020-00208-5>
- Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, 43(1), 93–114. <https://doi.org/10.1007/s11251-014-9328-3>
- Schuessler, K., Fischer, V., & Walpuski, M. (2024). Investigating construct validity of cognitive load measurement using single-item subjective rating scales. *Instructional Science*, 53(1), 71–97. <https://doi.org/10.1007/s11251-024-09692-6>
- Sundeen, T. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74–88. <https://doi.org/10.1016/J.ASW.2014.03.003>

- Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review*, 35(4), 95. <https://doi.org/10.1007/s10648-023-09817-2>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer. <https://doi.org/10.1007/978-3-319-24277-4>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 487662. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology*, 44, 1–11. <https://doi.org/10.1016/j.cedpsych.2015.11.003>
- Zhang, Y., Chen, B. L., Ge, J., Hung, C. Y., & Mei, L. (2019). When is the best time to use rubrics in flipped learning? A study on students' learning achievement, metacognitive awareness, and cognitive load. *Interactive Learning Environments*, 27(8), 1207–1221. <https://doi.org/10.1080/10494820.2018.1553187>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.